December 2010

Opinions and Editorials-Open and Accessible Supplemental Data: How Libraries Can Solve the Supplemental Arms Race

Elizabeth R. Lorbeer

University of Alabama at Birmingham, elorbeer@gmail.com

Heather Klusendorf

EBSCO Information Services, Hklusendorf@ebSCO.com

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>

Part of the [Library and Information Science Commons](#)

Recommended Citation

Lorbeer, Elizabeth R. and Klusendorf, Heather (2010) "Opinions and Editorials-Open and Accessible Supplemental Data: How Libraries Can Solve the Supplemental Arms Race," *Against the Grain*: Vol. 22: Iss. 6, Article 15.
DOI: <https://doi.org/10.7771/2380-176X.5693>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Op Ed — Open and Accessible Supplemental Data: How Librarians Can Solve the Supplemental Data Arms Race

by **Elizabeth R. Lorbeer** (Associate Director for Content Management, Lister Hill Library of the Health Sciences, University of Alabama at Birmingham) <lorbeer@uab.edu>

and **Heather Klusendorf** (Media Relations Coordinator, EBSCO Information Services)
<Hklusendorf@ebSCO.com>

Imagine a world where all research data belong to the community and are not subject to restriction or fee for use among the many, a world where the Internet advances sharing rather than creating new technologies for locking down data. In this world, the data is “open” and yet still protected, allowing all researchers to benefit from shared experimental data. Open data is described as “a philosophy and practice requiring that certain data be freely available to everyone, without restrictions from copyright, patents or other mechanisms of control.”¹ Researchers are increasingly sharing their experimental data on a global stage by making their bench work research accessible on research repository Websites. Creation of data drives new discovery and is the foundation of scholarly output in peer-reviewed journals. Open data allows for transparency, encourages debate and differential interpretations and is naturally allied with the Open Access movement in scholarly publishing. With increased pressure from the academic community and national government to make research freely assessable to the public, the Open Data movement strives to make the raw building blocks of knowledge widely available. Many researchers do not have access to platforms for housing and making this data available for future research, and publishers are beginning to cease the habit of housing this data; librarians may be the perfect custodians for managing supplemental data on a long-term basis in an open environment.

The practice of making the final product, the published paper, with all its supplemental attributes, easily findable as a whole is currently non-existent. While database providers offer indexing and abstracting of published literature, they do not offer external link outs to supplemental data unless identified by the publisher. This means that scholarly research is often read in its finished state organized by controlled subject headings without clues to how the organized research came to be. However, peer-reviewed publications require their authors to supply a link out to research data for public review. Authors who are funded by national agencies or are willing to share their research efforts

widely are tasked with footnoting their papers with information on where to find the supplemental materials, such as at their institution’s research repository (a publicly accessible Website) or requiring readers to request the materials through personal correspondence. Increasingly, though, a new trend is emerging where publishers are “ending the supplemental data arms race”² and no longer requiring authors to submit their supplemental data. This means that publishers will no longer house or demand that authors house this important building-block data.

Publishers are getting out of the business of accepting supplemental data to preserve scientific integrity and the peer review process. In an August 11, 2010 open letter, the Editor-in-Chief of the *Journal of Neuroscience* announced that the publication was no longer accepting supplementary materials with submitted manuscripts or hosting supplemental materials of accepted papers on its Website. The editor raised concerns that the current review practice and the depth of that review were questionable for supplemental materials.³ The responsibility for making supplemental materials available now resides fully with the author. Authors are instructed to “include a link to supplemental material on their own site” and leave it up to public opinion to comment on the validity of their data. But linking field research to data output from the published manuscript requires that the author is willing to share their data and be able to deposit in a repository that will insure long-term storage. Unless the author is mandated by a funding source to make his/her experimental data publicly accessible, many may choose to keep it locked in their laboratory notebook and unavailable for review.

What does this mean for librarians? It may mean an opportunity to preserve and protect this essential supplemental data. Our users regularly conduct disparate searches in bibliographic databases, search engines, and preprint servers to find all the pieces of discipline — specific data that relates to their work. As

daunting as that is, search engines and Web crawlers can be customized to pick through the Web to retrieve lab notes, data sets, podcasts, and accompanying material. As information providers, our efforts have tended to focus on supplying our users with content that the library has paid for by subscription or through an aggregate. However, as information providers we will need

to go further for those looking to account how qualitative and quantitative data was gathered. Even though methodology and the resulting sections of published papers remain critical, having access and being able to systematically examine the data allows for more transparency and genuine contributions to the discipline. As

more commercial and society publishers nix supplemental materials, a likely place to store these items will be on an institutional repository — perhaps one monitored by the library.

Librarians are best situated to preserve and curate data in their institution’s repository. An academic-sponsored repository provides a safe and organized place to permanently archive and share the results of research efforts. Librarians are naturally positioned to collect and assist in tagging metadata so it can be searched for and located. Issues of copyright, long-term preservation, and embargoed access will need to be tackled with local policy. For authors who deposit and point back to their home repository, the institution will need to have a policy in place for digital preservation to ensure accessibility. However, a larger debate within the preservation community will need to take place about ensuring access to data beyond its current digital state. For instance, as formats change, will digital bits be upgraded to the latest file format? How useful will supplemental data be years from now, and should only the published, peer-reviewed product be part of the cultural record? Not every researcher will want to share their proprietary data, and some may demand a toll be collected for access.

continued on page 43

R2: Converting Librarians to Cybrarians

The R2 Digital Library from Rittenhouse makes it easy for librarians everywhere to offer more than 1,000 health sciences titles online.

SAVE MONEY. Avoid “all-or-nothing” prices. Add more titles immediately with the push of a button.
SAVE TIME. Our easy-to-use interface meets needs of librarians. If you can surf the web, you can use R2.
MANAGE SMARTLY. Monitor usage statistic in real time, so you can make informed decisions.

Move into library management’s digital age with a name you can trust: Rittenhouse.

Rittenhouse
BOOK DISTRIBUTORS

DISTRIBUTING THE LEADING SCIENTIFIC, TECHNICAL AND MEDICAL PUBLISHERS

511 Feheley Dr., King of Prussia, PA 19406 • 800-345-6425 • www.r2library.com • www.rittenhouse.com

Op Ed from page 42

Undoubtedly, this will place pressure on database and Web discovery providers to manufacture online tools to draw attention to linkouts to supplemental materials that exist beyond the confines of the controlled search environment. Right now, some providers offer value-added features on their sites to expose data within their depths, such as Elsevier’s SciVerse platform. Retrieving results outside of the provider’s site will require user-designed engines to crawl content. Historically, publishers and vendors have looked unfavorably upon Web crawlers due to additional stress that such traffic can place on a system, temporarily shutting down a school’s Internet Protocol address to cease crawling activity. Can standards

change this, forcing publishers to create sites that allow content to be crawled so users can unearth supplemental materials that reside within the publisher’s online environment?

The burden remains for libraries, publishers and online indexing services to be able to point the public to supporting data that produced the published manuscript. How data is being shared, especially among research communities, may require a significant change to long-standing practices. As authors willingly share their scholarly output, and make their research more visible, they must also guard their rights on how the data can be used. Academic and research centers pay careful attention to how their data is exposed to deter any lost income from potential inventions. Librarians can be the gatekeepers who can help to preserve, protect, and make available supplemental data,

creating an increasingly open research environment where sharing rather than locking data is the norm. 🌱

Endnotes

1. Open science data. Wikipedia. Accessed October 18, 2010. http://en.wikipedia.org/wiki/Open_data
2. Davis, Phil. Ending the Supplemental Data “Arms Race” on The Scholarly Kitchen. Accessed October 18, 2010. <http://scholarlykitchen.sspnet.org/2010/08/16/ending-the-supplemental-data-arms-race/>
3. Announcement Regarding Supplemental Material. Accessed October 18, 2010. <http://www.jneurosci.org/cgi/content/full/30/32/10599>

Rumors from page 40

gling to maximize profits from this fast-growing digital sector. Inefficient print production processes are hampering scalable, affordable digital output. c) Only 7% of publishers are implementing enhancements to their eBooks. Suggesting that most publishers are not aware of the EPUB standard’s inherent support for content enhancement, including audio and video.

d) 61% of trade/consumer publishers support the EPUB standard. 18% more than any other publisher type. e) 64% of all publishers are offering titles in eBook format. Up 11% from the first survey. To find out full results of this survey: <http://www.aptaracorp.com/ebook-survey2/?elq=b7dca85d436544b8a1ed469c3c163609>

So – you may have noticed their schizophrenic nature or these Rumors – between print and e! Reminds me of the 2002 Charleston Conference theme – TWO FACES HAVE

I: One for Books and One for Bytes. Ain’t it the truth?

Continuing with my schizophrenia – You know – **Books are not just words**, groups of words or sections of words. Books are artifacts of our civilization, memories, communication tools. One of the speakers at the Charleston Conference (Allen Renear) said: “There is too much to read; text mining and strategic reading are necessary. We need disaggregated

continued on page 51